Title: Lineage maintenance due to assortative mating but not ecological divergence in a sympatric sibling species pair

Mitchell Irvine¹*, Zachary Stewart¹, Nagalingam Kumaran², Chapa G. Manawaduge¹,
Solomon Balagawi³, Brendan Missenden⁴, Melissa Starkie⁴, Anthony R. Clarke¹, David Hurwood¹, and Peter Prentis¹

* Correspondence: mitchell.irvine@hdr.qut.edu.au

Mitchell Irvine (ORCID ID – 0009-0004-2338-8463)

¹School of Biological and Environmental Sciences, Queensland University of Technology,
GPO Box 2434, Brisbane, Qld, Australia
Abstract

Background

Lineage theory suggests that the development of traits associated with reproductive isolation coupled with traits associated with ecological differentiation are essential for the maintenance of sympatric lineages. However, the relative importance of these factors have rarely been tested simultaneously. *Bactrocera tryoni* and *B. neohumeralis* are a tephritid fruit fly species pair that have significant overlap in geographic range and host use, with time of male mating the only known difference in their mating systems. Using this system, we tested the relative importance of ecological differentiation versus assortative mating in sympatric lineage maintenance.

Results

Genome-wide SNP analyses found strong genetic differentiation between the species with no evidence for hybridization in the field. Most outlier SNPs were restricted to narrow regions towards the centromeres and telomeres of chromosomes. Enrichment of annotation terms indicated an overabundance of genes with the ‘abnormal neuroanatomy’ term. Terms of interest associated with sleep and circadian rhythm, potentially important to the allochronic reproductive barrier, were non-enriched. Ecological data found no evidence for ecological divergence or competitive displacement between the two species based on significant positive correlations between species numbers trapped at different times of the year, trapped in different habitats within a region, or when reared from fruit.

Conclusions

Our study highlights the significance of assortative mating over ecological differentiation for sympatric lineage maintenance of the *B. tryoni/B. neohumeralis* sibling pair. The paper
represents one of the most well-characterized examples of the importance of genomic
divergence in the coexistence of two closely related species.

**Keywords 3-10 words**

*Bactrocera tryoni, Bactrocera neohumeralis*, lineage maintenance, coexistence, Tephritidae,
competitive displacement, recombination

**Background**

Species are increasingly defined as separately evolving lineages [1-4]. If lineages remain
separate, then attributes used to define a species, such as unique mate recognition, distinct
phenotypes and adaptation to specific habitats, may (or may not) emerge through selection,
genetic drift, or pleiotropic change [5, 6]. However, while lineage theory may be coming
widely accepted, the questions faced by evolutionary biologists of what drives lineage
divergence and then keeps lineages separate in sympathy, remain [7, 8].

Central to our understanding of lineage maintenance, particularly between sympatric or
parapatric lineages, are the opposing roles of natural selection and recombination on the
formation or dissolution of lineage boundaries. Natural selection promotes the development
of lineage boundaries through the formation of locally adapted gene complexes; while free
recombination acts to dissolve them through breaking down the associations of such
complexes [9-11]. A reduction in recombination rates is considered necessary for lineage
boundaries to persist and strengthen over time [12]. Traits associated with reproductive
isolation, such as those involved in assortative mating, are considered effective mechanisms
for recombination reduction and essential for lineage divergence in the presence of gene flow
when coupled with the development of traits associated with local adaptation [10, 13].

On an ecological scale, for two lineages to persist in sympathy is considered to be reliant on
two criteria; their ability to overcome the effects of hybridization through the development of
reproductive isolation, and the development of niche partitioning to avoid inter-lineage competition asymmetries[11, 14, 15]. Without meeting these criteria, theory suggests that lineages will likely dissolve through hybridization or cease to exist in sympatry through competitive displacement [14, 16]. Few empirical studies have examined evidence for assortative mating and competitive displacement simultaneously in the same species pair. The Australian frugivorous tephritids (Diptera: Tephritidae), *Bactrocera tryoni* (Froggatt) and *Bactrocera neohumeralis* (Hardy), are a fruit fly sibling pair which have been studied in terms of their speciation histories and genetic stability for over 60 years [17-26]. Genetically, *B. tryoni* and *B. neohumeralis* are very similar [23, 27, 28], with the sharing of polymorphisms in coding and non-coding regions indicating continued genetic exchange or recent separation [22, 24, 27]. The species show only minor variation in their morphology; the humeral calli is bright yellow in *B. tryoni* and brown in *B. neohumeralis* [29]. While intermediate colour variants are well documented, genetic analysis using microsatellites suggest that intermediates represent phenotypic variation within either species rather than hybrid individuals [21, 24, 30, 31]. While *B. tryoni* is typically more abundant than *B. neohumeralis*, the geographic ranges of the two species extensively overlap, with the endemic range of *B. neohumeralis* being entirely encompassed by the range of *B. tryoni* except for a few islands in the Torres Strait north of Queensland (Fig. 1a,b) [32]. Both *B. tryoni* and *B. neohumeralis* are highly polyphagous, using fruit from multiple plant families for breeding, with host utilization data [presented as ‘simple’ lists of host fruit that either species have been reared from at least once with little or no further information] showing significant overlap between the pair (Fig. 1c) [33-35].
Figure 1. Range and host overlap between Bactrocera tryoni and B. neohumeralis. A) Distribution of both species in Australia. B) Distribution of both species in the Cape York Peninsula and Torres Strait Islands. C) Host usage overlap between species. Numbers in circles represent the known number of host species utilized. Range and host data gathered from Hancock, E.L. [33], Leblanc, Vueti [34], QDPC [35]. Note: Recent expansion of B. tryoni into central and western Victoria is not shown because it is still under regulatory control in many places and is not associated with the species’ evolutionary history. So closely related are B. tryoni and B. neohumeralis that only one known feature unambiguously defines them as separate biological species: their time of mating [19, 36, 37].
Bactrocera neohumeralis mates during the middle of the day at high light intensity over a mating window of three to seven hours, while *B. tryoni* initiates mating during dusk at low light intensity in a mating window of approximately one hour [38]. This difference in mating time is the only known mechanism maintaining reproductive isolation between the species [18-20, 26, 37], with other components of their mating system not known to be different [38-40], including an absence of post-zygotic barriers [41]. Recent work shows that female *B. neohumeralis* will, like *B. tryoni*, also mate at dusk, at which time *B. tryoni* males do not discriminate between females of either species [41]. This caused Yeap et al [41] to conclude “the mating time difference may be a weaker reproductive isolating barrier than once assumed”.

If mating time is a weak reproductive barrier between the two species, then other factors must be playing a role in maintaining separate lineages; for example, genomic incompatibility [42, 43]. Additionally, for herbivorous insects such as fruit flies, lineage divergence is largely considered to be driven by host switching and host specialization that reduces the effects of competition and natural enemies [44-48] and so ecological factors also need to be considered when investigating lineage maintenance. Conversely the lineages may not be evolutionarily stable, in which case evidence for genetic recombination should be expected from sympatric field collections of the two taxa.

In this study we conducted a comprehensive investigation of how the *B. tryoni* and *B. neohumeralis* lineages have been maintained in sympatry through dense geographic sampling of nuclear genomes and rigorously collected ecological data. Furthermore, we determined whether specific genomic regions are associated with reproductive isolation between the two species. Despite utilizing the same resources in their environment, we found no evidence of competitive displacement between the two species. Notably, our findings demonstrated that *B. tryoni* and *B. neohumeralis* exhibit strong genomic differentiation that can be attributed to...
specific genomic regions that likely contribute to maintaining the two as distinct lineages in sympathy.

Results

Molecular Data

Read processing, alignment and SNP calling

After demultiplexing samples, we obtained between 2,596,430 and 5,228,647 high-quality sequence reads for each sample. Alignment to the *B. tryoni* genome assembly using BWA-MEM indicated that most samples had high mapping rates (median alignment percentage = 74.98%), with a small number of samples having low sequence alignment (minimum alignment percentage = 0.38%); see Table S2 and S3 for full details of read counts and alignment percentages. 28,334 single nucleotide polymorphisms (SNPs) were identified using Freebayes from the two species that met our threshold for inclusion.

Populations cluster by species and not geography

We found that populations clustered based on species and not geography in all analyses. Principal Component Analysis (PCA) (Fig. 2) and fastSTRUCTURE (Fig. 3) both resolved species-specific groups, demonstrating clear genetic delineation of the two species and an absence of intermediate genotypes between the species. Several *B. neohumeralis* samples (genotype codes 526, 528, 529, 530, and 531 from the NS location) appear to be outliers based on PCA (See Figure S1), but one sample from the same location (527) grouped with the majority of *B. neohumeralis* samples in ordination space. Additionally, the chooseK.py function of fastSTRUCTURE predicted three population clusters, with the outlier samples from the PCA representing the third cluster (See Fig. S2). These outlier samples had poor genotype calling rates (min = 27.9%, max = 32.3%) compared to sample 527 (91.3%) and most other samples (median = 89.1%). As such, we believe these to be outliers to result from sequencing issues and were excluded from further analysis.
Patterns of differentiation among populations (FST) were also lower between populations within a species (B. neohumeralis = 0.0025796 when ignoring outlier samples, B. tryoni = 0.0036269) and two orders of magnitude higher between species (0.10843). Identity-By-State (IBS) analysis also provided insight into the genetic homogeneity of the two species, with B. tryoni having a mean value of 0.797 and B. neohumeralis having a mean value of 0.790 (species with a value closer to 1 are more identical) suggesting that both species are approximately equally homogenous.

**Figure 2.** Principal component analysis (PCA) of single nucleotide polymorphisms (SNPs) for B. neohumeralis and B. tryoni. Values from the first two eigenvectors plotted for each sample. Eigenvectors one (horizontal) and two (vertical) explain 10.204% and 1.68% of the variance in the data, respectively.
**Figure 3. fastSTRUCTURE analysis for *B. neohumeralis* and *B. tryoni*.** Results at cluster size (K) equals two for single nucleotide polymorphisms (SNPs) for both species grouped by collection site. Red and blue colouration of the bars indicates the inferred population admixture proportions as a percentage (numbers not shown). Three-digit numbers along the bottoms of each row of bars are genotype ID codes. The intermediate grouping includes samples with an intermediate phenotype. All three intermediate samples (genotypes 803, 805 and 807) presented as *B. neohumeralis* genotypes.

*Outlier loci are clustered in specific chromosome segments between the species*  
Using outlier detection in BayeScan, we found 357 of the 28,334 SNPs (1.26%) displayed a significantly elevated level of differentiation between the species. The distribution of outlier SNPs along autosomal chromosomes (Fig. 4, NC_ prefix) was largely restricted to ~ 16 narrow genomic regions. These narrow genomic regions with outlier loci were not uniformly distributed across the five autosomal chromosomes with 10 occurring on chromosome ends.
and two near the centre of chromosomes. A relatively small number of outlier SNPs were
mapped to ‘unplaced contigs’ (See Fig. S3).

![Outlier plot representation of SNPs for B. neohumeralis and B. tryoni](image)

**Figure 4. Outlier plot representation of SNPs for B. neohumeralis and B. tryoni** identified on the five autosomal chromosomes. $F_{ST}$ values for each SNP were computed and are plotted as the Y-axis value (left) for each chromosome, with the X-axis representing the position in base pairs along each chromosome (scaled proportionately); these chromosomes and contigs are shown as separate rows in the figure, with their NCBI sequence identifier shown to the right. Outlier SNPs are highlighted in blue.

We investigated genes that contained outlier SNPs or were in close proximity to outlier SNPs, resulting in the identification of 266 genes. Enrichment of annotation terms indicates an overabundance of genes with the ‘abnormal neuroanatomy’ term (n=11 genes, p= 0.0164) within these candidate genes. Several neurology-related terms which were not enriched but still of interest included ‘nervous system development’, ‘central nervous system development’, ‘sensory system’, and ‘sensory organ development’. Additional non-enriched terms of interest included sleep and circadian rhythm-related terms, including: ‘abnormal sleep’, ‘circadian rhythm’, ‘abnormal circadian rhythm’, and ‘abnormal circadian behaviour’. Inositol-related terms were statistically enriched, which included ‘inositol biosynthetic
process’ (n=1 gene, p= 0.0115) and ‘inositol-1,4,5-trisphosphate 3-kinase activity’ (n=1 gene, p= 0.0467). The 19 genes annotated with these terms are shown in Table S4.

Ecological Data

*No differences in seasonal activity between the species*

Trap data collected over several years showed no difference in seasonal activity between *B. neohumeralis* and *B. tryoni* (Fig. 5 & Fig. S4). For any given sampling event there were generally significantly more *B. tryoni* than *B. neohumeralis* trapped (See Table S5), but the abundance of both species over time were highly linearly correlated across nine sites with correlation values ranging from: \( r = 0.776 \) (Ayr, \( p < 0.001 \)) to \( r = 0.993 \) (Maryborough, \( p < 0.001 \)) (See Table S6). South Johnstone was the only site with a non-significant result (\( r = 0.926; p = 0.074 \)) however, limited collections were recorded at this site (\( n = 4 \)).
Figure 5. Correlation of seasonal abundance of *B. tryoni* and *B. neohumeralis*. Samples were collected at multiple locations throughout Queensland, Australia. Scatterplots were developed using monthly mean of trap catches over 2 - 7 years. Data from May [49].

No differences in habitat use between the species

There was a significant correlation between *B. tryoni* and *B. neohumeralis* abundances in their habitat use (Fig. 6). Within a habitat type, there was a significant correlation in the two species monthly abundances for all habitat types, with the correlation ranging from *r* = 0.54 (dry forest, *p* < 0.001) to *r* = 0.90 (mixed farming, *p* < 0.001) (See Table S8).
**Figure 6. Correlation in habitat use of Bactrocera tryoni and B. neohumeralis.** Samples were collected from different human-defined habitat types within the Bundaberg district of Queensland. Each data point is the adult trap catch of both species for a trap run for three consecutive days. For each habitat type, nine traps were run per month for seven months.

**Limited differences in host use between species**

One-thousand, six hundred and seven fruits were sampled from the field and the emergence of B. tryoni and B. neohumeralis adults counted. For white sapote, mulberry, peach, plum, carambola and nectarine, sufficient samples were collected to carry out formal analysis. Highly significant correlations between the number of B. tryoni and B. neohumeralis adults reared from each of these fruit types was found (correlation ranging from $r = 0.904$, $p < 0.001$ for white sapote to $r = 0.952$, $p < 0.001$ for peach) (Fig. 7 and Table S9). For guava, two collections occurred: a single B. tryoni and no B. neohumeralis emerged from the September
2013 collection while 1609 *B. tryoni* and 1353 *B. neohumeralis* emerged from the October
2013 collection; for feijoa a single collection occurred with equal numbers [71 flies each] of
*B. tryoni* and *B. neohumeralis* emerging. For sapodilla, *Syzygium* spp, grumichama, hog plum,
jabotica, white mulberry, avocado, black sapote, longan, mango and cashew, *B. tryoni* but no
*B. neohumeralis* emerged.

![Graphs showing correlation in host use of *Bactrocera tryoni* and *B. neohumeralis*.](image)

**Figure 7.** Correlation in host use of *Bactrocera tryoni* and *B. neohumeralis*. Each data
point is the number of adult flies reared from an individual collection of multiple fruit pieces
from a research orchard at Nambour, Queensland.

**Discussion**

*Assortative mating and reduced recombination in the absence of local niche adaptation*
Genetic comparisons of *B. tryoni* and *B. neohumeralis* individuals collected throughout the overlapping range showed no evidence of hybridization between the species. This was despite samples of each species being collected from the same traps, i.e., in immediate sympatry. Several intermediate phenotypes were collected, however, these samples presented strongly as *B. neohumeralis* genotypes with no indication of hybridization. This is consistent with the findings of prior studies of the species pair and supports the theory that strong reproductive barriers exist in non-laboratory conditions [21, 24, 41]. While Yeap et al. [41] suggest the allochronic reproductive barrier between these taxa may not be as strong as initially proposed due to their ability to produce viable hybrid offspring under laboratory conditions, the lack of hybrid individuals identified from natural populations in this study suggest that assortative mating is effective in maintaining species boundaries.

The linkage of genes involved in assortative mating with those associated with habitat preference and performance is considered necessary for sympatric speciation [50, 51] and has been revealed to have involvement in host switching and specialization of phytophagous insects such as apple maggot flies and pea aphids [52, 53]. In these studies, the genes involved appear to map tightly together along narrow regions of the chromosome that is consistent with the patterns observed in our data. These tightly mapped genes have been theorized as a mechanism for speciation in the presence of gene flow and are generally considered to occur through chromosomal rearrangements, such as inversions, where the genes are brought into physical linkage [50, 51]. A small region of modifier alleles can also act to reduce recombination between populations in the absence of chromosomal arrangements [12, 13]. Regardless of the mechanism causing genetic variation, the regions of differentiation are expected to involve locally adapted alleles or gene complexes and result in reduced recombination between populations [11, 13]. In our data however, there is an apparent lack of genes relating to local adaptation in the narrow regions containing outlier
SNPs. Instead, genes related to mating behaviour and circadian rhythm have been identified that we propose are responsible for separating the two species and forming the time-of-day premating reproductive barrier through assortative mating. While ecological divergence is generally required for sympatric speciation, sexual selection alone may be sufficient in forming species boundaries [54, 55].

In the absence of locally adapted genes/gene complexes under selection pressure, it is likely that the positioning of these narrow regions of genetic variation along the chromosome had the greatest influence on reducing recombination. Most outlier SNPs observed in our data were largely restricted to narrow regions located towards the telomeres or centromeres of chromosomes (Fig. 4). Reduced recombination has been observed in these regions in several taxa, including Drosophila, mice, rabbits, and humans [50, 56-60]. The combination of genes associated with assortative mating, and their chromosomal positioning may have sufficiently suppressed recombination, preventing gene flow between species without the need for selection pressure on locally adapted genes.

A lack of competitive displacement still allows co-existence

Competitive displacement has long been suggested as an important mechanism facilitating divergence and speciation [61-63]. The ability of two nascent lineages to coexist by minimizing competition, or by utilizing non-competitive ecological space, should be considered central to early lineage divergence and persistence [16]. Where two herbivore populations share resources, then displacement can occur in time (i.e. seasonal activity), location or host usage [64, 65]. However, our data shows no evidence for this in the B. tryoni/B. neohumeralis sibling pair. Both species were active at the same time of the year and in the same habitats within a landscape. Bactrocera tryoni was recovered from a greater range of host fruits than B. neohumeralis and that may be seen as evidence for competitive
displacement (i.e. competitive advantage by *B. neohumeralis* causing *B. tryoni* to use novel hosts), but we consider it more likely that this pattern simply reflects *B. tryoni*’s greater abundance in the environment and so greater likelihood of being sampled, rather than a competition avoidance mechanism. This is because when both species were reared from a fruit, their numbers were closely correlated. In cases where competition between fruit flies has been known to cause changes in field host use patterns, one species will dominate in a given host, pushing a second species to host fruits that the first only irregularly uses [66, 67]: we did not observe this pattern. The positive correlation in the number of emerging adults of both species from fruit also offers field data supporting the work of Kay [68], who in laboratory studies found no interspecific larval competition between the species. Our data suggests that, contrary to classical niche theory, ecology has little influence on speciation and lineage maintenance in our sibling pair and that lineage maintenance is largely driven by genetic factors and assortative mating.

Our study provides a comprehensive comparison of both molecular and ecological data, between two sympatric sister taxa. High levels of niche conservatism and a lack of competitive displacement suggests that the formation and persistence of these taxa occurred in the absence of ecological selection pressures. Further, the development of assortative mating and the reduction of recombination of tightly linked sites within chromosomal telomeres and centromeres appears to have sufficiently reduced geneflow between the species to maintain the lineage boundaries in sympatry, suggesting that chromosomal architecture and reproductive isolation in the form of assortative mating, had a greater influence on lineage divergence than ecological factors.

**Methods**

*Molecular data sample collection*
Individuals of *B. tryoni* and *B. neohumeralis* were collected in December 2020 and December 2021, using cue lure baited Steiner [69], Lynfield [70] or Paton traps [71]. These traps were deployed by Queensland Department of Agriculture and Fisheries (QDAF) as part of ongoing Biosecurity Queensland (BQ) surveillance. A subsample of the trapping sites were selected from the existing network to represent the geographic range of both species (Fig. 1). Up to 10 individuals of each species were collected per site with a total of 81 *B. tryoni* and 105 *B. neohumeralis* individuals collected from 19 sites. Detailed trapping records are included in Table S1. Identity of the samples were confirmed following the descriptions in fruit fly identification handbooks [29, 72], and the samples were stored at -20°C in 100% ethanol.

**Extraction and sequencing**

DNA was extracted using the QIAGEN DNeasy® Blood & Tissue Kit following the manufacturer’s protocol. Samples were screened for quality and quantity, using both gel electrophoresis and Qubit assay, and then sent to Diversity Arrays Technology Pty Ltd, Canberra (DArT P/L), for DArTseq high-density genotyping. A PstI/SphI, restriction enzyme combination was used by DArT P/L, and the fragments (up to 90bp) were sequenced on an Illumina Hiseq2500 as single end reads.

**Data sourcing**

All relevant analyses made use of the *B. tryoni* genome assembly (Genbank accession = GCF_016617805.1).

**Read alignment and SNP calling**

Reads were demultiplexed using process_radtags as part of Stacks v2.60 [73]. Read mapping took place using BWA-MEM [74] to produce a SAM alignment file with appropriate read group specification for each sample. Subsequently, samtools [75] produced sorted and indexed BAM alignment files. Freebayes v1.3.6 [76] was used in an iterative process to predict SNPs in all samples. Firstly, each sample had SNP calling performed individually.
using freebayes; output Variant Call Format (VCF) files were left-aligned using BCFtools [77] and block substitutions were decomposed using vt [78]. Resulting VCF files were merged using BCFtools, and this VCF file was used as input for a second round of SNP calling using freebayes (-@ file.vcf --only-use-input-alleles) to ensure all samples were genotyped at the same locations. The variant calls were filtered using a method based upon James, Arenas-Castro [79] wherein SNPs were removed if they had missing data in > 50% of the population, had a quality score < 30, a depth < 3 read alignments, and a minor allele count < 1; we also filtered SNPs with a minor allele frequency < 5%.

Population structure

To investigate whether admixture was occurring between the two species, we used fastSTRUCTURE v1.0 [80]. A simple prior of K=1-10 was used, after which the chooseK.py utility program provided with fastSTRUCTURE was used to assess the optimal number of genetic clusters to explain the structure in the species populations. Plots were generated using pophelper v2.3.1 [81] in R.

Additionally, we ran PCA to provide further visualisation of the variations within and between the two species, using the SNPRelate v1.24.0 [82] package in R. The linkage disequilibrium-based SNP pruning function provided with this software was run with default parameters to mitigate any bias associated with linked SNP clusters upon the downstream PCA. We similarly performed this SNP pruning prior to calculation of IBS proportions for samples using SNPRelate.

Mean FST values [83] were calculated both within and between the two species using VCFtools [84]. Within species comparisons were conducted for each species separately and samples from each collection site were considered as separate populations; for between species comparison, the samples were grouped by species and entered as two separate populations.
Outlier SNPs and gene candidates

Variant predictions were converted from VCF to GESTE format using PGDSpider v2.1.1.5 [85], then outlier SNPs were predicted using Bayescan v2.1 [86] with default parameters. Genes containing outlier SNPs within their gene model (exons and/or introns), as well as genes who were the nearest to an intergenic outlier SNP were identified using a custom script.

Candidate genes were queried against the *Drosophila melanogaster* gene models [87](FlyBase release FB2023_01) using MMseqs2, and the FlyBase automated gene summaries of the best hit were parsed for functional and phenotypic annotations which we attributed to our sequences if a minimum E-value of $1 \times 10^{-10}$ was achieved.

Seasonal abundance

A historical dataset [49] was used for the comparison of seasonal abundance of *B. neohumeralis* with *B. tryoni*. This dataset has been previously used to report *B. tryoni* [but not *B. neohumeralis*] seasonal phenology [88], and so the phenology patterns *per se*, are not explored here. Rather, we focus on evidence for temporal niche segregation in *B. neohumeralis* and *B. tryoni* through paired t-tests and Pearson correlation analysis using the R package ggpubr [89]. We predict that if there is temporal niche displacement between the two species, then there should be no, or even negative correlation between their seasonal abundances. We interpret significant positive correlation between their populations as a lack of evidence of temporal niche displacement.

Detailed methodology on the fly trapping dataset is described elsewhere [49, 88]. In summary, 15 trapping sites across Queensland, Australia, were used, ranging from temperate Stanthorpe in the south to tropical Cairns, ~1500km to the north (See Table S5). Fly trapping was carried out for up to six years per location using a fruit-based liquid lure trap [90]. At
each location ten traps were maintained, with traps cleared weekly. The data in May [49] varies between sites in how it is presented, and modifications were needed in order to be able to directly compare *B. tryoni* and *B. neohumeralis* populations and six sites were excluded as seasonal data were not available (see Table S5 for details). The t-tests and correlation analysis between *B. tryoni* and *B. neohumeralis* populations was conducted on a site-by-site basis (i.e., data were never collated across sites), so the different manipulations performed on the site-specific data sets did not affect results.

**Habitat use**

This study compared abundance of *B. tryoni* and *B. neohumeralis* across different human-defined habitat types to determine habitat use patterns. Six habitat types, grassland, suburbia, horticultural farming, mixed sugarcane and other crops farming (mixed farming hereafter), dry forest, and wet forest were chosen for the study in the Bundaberg region of Queensland (~350 km north of Brisbane). Specific descriptions of each habitat type are outlined in Table S7. For each habitat type there were nine replicate fruit fly traps. Trapping commenced in September 2010 (spring) and was completed in March 2011 using cuelure baited modified Steiner traps [69]. The trapping period covers the period of peak yearly activity for *B. tryoni* [91]. The trapping program involved setting up the traps on a monthly basis at each trap site, removing the traps at each site three days after setting up and replacing the traps with fresh wicks and attractants in the following month. Flies collected over the three days in each month were counted and identified to species level using a fruit fly identification guide [29]. Paired t-tests and Pearson correlation analysis on the total trapping dataset for a habitat type was performed to compare the relationship in abundance of the two different species at that habitat type.

**Host fruit use**
Fruits were collected from the QDAF Maroochy Research Facility, Nambour, Queensland (~100 km north of Brisbane) from separate orchard blocks that consisted of either mixed tropical fruit varieties, stone fruit or mangoes, as well as some isolated trees. Small numbers of ripe or mature green fruit were sampled on a weekly or fortnightly basis between December 2012 and October 2015. Fruit were selected from a number of trees within the orchard block and picked from various heights and aspects to obtain a random sample. Fruits collected were: white sapote (**Casimiroa edulis**), mulberry (**Morus nigra**), peach (**Prunus persica**), plum (**Prunus domestica**), carambola (**Averrhoa carambola**), nectarine (**Prunus persica**), guava (**Psidium guajava**), sapodilla (**Manilkara zapota**), **Syzygium** spp, grumichama (**Eugenia brasiliensis**), hog plum (**Spondias mombin**), jabotica (**Plinia cauliflora**), white mulberry (**Morus alba**), avocado (**Persea americana**), black sapote (**Diospyros nigra**), longan (**Dimocarpus longan**), mango (**Mangifera indica**) and cashew (**Anacardium occidentale**).

Fruit were placed in paper bags and transported to the QDAF Brisbane laboratories. Fruit were counted and weighed and then placed on gauzed plastic containers over vermiculite, in plastic boxes with gauzed lids to allow ventilation. Boxes were held in a Controlled Environment Room (26°C and 70%RH) to allow insects to develop through to the pupal stage. The vermiculite was sieved weekly until all insects had exited fruit and pupated, and then fruit was inspected before being discarded. Fruit fly pupae that were reared out of fruit samples were placed into small plastic boxes with gauzed lids containing vermiculite. Once adult fruit flies had emerged from pupation, they were identified to species level using morphological taxonomic characters. The abundance of the two fly species from a host was compared using a paired t-test and Pearson correlation analysis.

**List of abbreviations**

BQ – Biosecurity Queensland
IBS – Identity by State

PCA – Principal Component Analysis

QDAF – Queensland Department of Agriculture and Fisheries

SNP – Single Nucleotide Polymorphism

VCF – Variant Call Format

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Competing interests

The authors declare that they have no competing interests.

Funding

This research was partially funded by the ARC ITTC project IC150100026, Centre for Fruit Fly Biosecurity Innovation awarded to A.R.C. and P.P.

Authors’ contributions
A.R.C. and P.P. conceived the study. N.K., C.G.M., S.B., B.M. and, L.S. conducted field work and collected samples for molecular and ecological analysis. M.I., N.K., A.R.C. and, P.P. designed, collated, and analyzed ecological data. Z.S. and P.P. designed and performed bioinformatics and SNP analysis. M.I., Z.S., P.P. collated analyzed molecular data. M.I., Z.S., N.K., C.G.M., M.S., A.R.C., D.H. and P.P. contributed to the writing of the manuscript. All authors have read and approved the final version.

Acknowledgements

We thank J. Ryan for assisting with generating the distribution map, E. Vueti who carried out much of the Bundaberg sampling as part of an uncompleted postgraduate study, and L. Senior (QDAF) for her assistance with rearing flies from fruit that provided data for the host utilization study. We also thank The Southern Downs Regional Council, The Granite Belt Growers Association, B. Doohan, The Greenhouse, Lockhart River, D. Masters, J. Kerr and J. Whistler for their assistance in collecting samples for the molecular analysis.

Author details

1School of Biology and Environmental Science, Queensland University of Technology, GPO Box 2434, Brisbane, Queensland, Australia. 2Commonwealth Scientific and Industrial Research Organization (CSIRO), Ecosciences Precinct, Dutton Park, Queensland, Australia. 3Horticulture and Forestry Science, Department of Agriculture and Fisheries, P.O. Box 652, Cairns, Queensland, Australia. 4Department of Agriculture and Fisheries, Ecosciences Precinct, Dutton Park, Queensland, Australia

References


50. Ortíz-barrientos D, Reiland J, Hey J, Noor MA. Recombination and the divergence of

51. Yeaman S. Genomic rearrangements and the evolution of clusters of locally adaptive

Allopatric genetic origins for sympatric host-plant shifts and race formation in Rhagoletis.

53. Hawthorne DJ, Via S. Genetic linkage of ecological specialization and reproductive

54. Servedio MR, Boughman JW. The role of sexual selection in local adaptation and

Sister species diverge in modality-specific courtship signal form and function. Ecology and

56. Carneiro M, Ferrand N, Nachman MW. Recombination and speciation: loci near
centromeres are more differentiated than loci near telomeres between subspecies of the

57. Carpenter AT. Synaptonemal complex and recombination nodules in wild-type

58. Payseur BA, Nachman MW. Microsatellite variation and recombination rate in the

59. Nachman MW, Churchill GA. Heterogeneity in rates of recombination across the


68. Kay BJ. The complexity of competition between three native Australian fruit fly species, Bactrocera tryoni, B. neohumeralis and B. jarvisi in a changing environment [Thesis (PhD)]: Queensland University of Technology; 2023.


72. Plant Health Australia. The Australian handbook for the identification of Fruit Flies. 3.0 ed. Canberra, ACT: Plant Health Australia; 2018.


